

Multiple Linear Regression Analysis of Factors Affecting Grain Yield

Ranran Jia, Xu Han

School of Mathematics and Computation Science, Anqing Normal University, Anqing Anhui
Email: 94633351@qq.com

Received: May 20th, 2019; accepted: Jun. 7th, 2019; published: Jun. 14th, 2019

Abstract

This paper mainly introduces how to use multivariate linear regression method to analyze the influencing factors of grain yield, and use IBM SPSS Statistics software to carry out analysis, and put forward relevant suggestions based on the analysis results.

Keywords

Grain Yield Analysis, Multiple Linear Regression, Significance

粮食产量影响因素的多元线性回归分析

郝然然, 韩旭

安庆师范大学数学与计算科学学院, 安徽 安庆
Email: 94633351@qq.com

收稿日期: 2019年5月20日; 录用日期: 2019年6月7日; 发布日期: 2019年6月14日

摘要

本文主要介绍了如何使用多元线性回归方法对粮食产量的影响因素进行分析和, 使用IBM SPSS Statistics软件对求出的结果进行分析, 并对分析结果结合实际提出相关建议。

关键词

粮食产量分析, 多元线性回归, 显著性



1. 引言

2018年中共中央、国务院印发了《乡村振兴战略规划(2018-2022年)》，其中要求贯彻落实加快农业现代化步伐，提升农业科技创新水平。农业也是国家发展的基础[1]，粮食产量更是我国经济发展不可缺少的指标。相关资料显示，国家近几年来粮食产量在不断提高，而粮食产量与哪些因素有关，关系如何等问题便成了要进一步提高粮食产量必须要解决的问题。

多元线性回归[2]是统计学中常用的分析方法之一，主要用于分析因变量与多个自变量之间的具体关系。一些文献中分析了多元线性回归与其他分析方法的区别等，并举出实例。本文根据近十年数据对粮食产量的影响因素进行分析，求出相关的回归方程，继而提出合理性建议。

2. 多元线性回归相关理论

2.1. 多元线性回归方程

多元线性回归分析是研究多个自变量与一个因变量之间的非确定关系的一种方法，其基础的方程结构为：

$$f(x_t) = a_0 + a_1x_{1t} + \dots + a_nx_{nt} + \mu_t$$

其中，因变量 $f(x_t)$ 为 $f(x)$ 的第 t 次的取值，自变量 x_{it} 为第 i 个变量 x_i 的第 t 的取值， a_i 为回归方程的回归系数， $(i \in (1, 2, \dots, n))$ ， μ_t 表示第 t 次的误差且满足 $\mu_t \sim N(0, \sigma^2)$ 。若 $\mu_t = 0$ ，则称 $f(x_t)$ 为理论回归方程。

在多元线性回归方程中，模型必须满足的条件为：

- 1) 自变量对因变量有显著的影响，并呈现出非确定关系(线性关系)；
- 2) 样本容量 > 回归系数的个数；
- 3) 残差服从正态分布。

2.2. 回归系数估计(OLS)

残差平方和公式为：

$$Q(\beta_0, \beta_1, \dots, \beta_n) = \sum (y_t - \beta_0 - \beta_1x_{1t} - \dots - \beta_nx_{nt})^2$$

即，若 Q 取最小值，则 Q 对 $\beta_0, \beta_1, \dots, \beta_n$ 的偏导数为 0。下将 Q 对 $\beta_0, \beta_1, \dots, \beta_n$ 求偏导数，并令偏导数为 0，可得到如下回归方程组：

$$\begin{aligned} n\beta_0 + \beta_1 \sum x_{1t} + \dots + \beta_n \sum x_{nt} &= \sum y_t \\ \beta_0 \sum x_{1t} + \beta_1 \sum x_{1t}^2 + \dots + \beta_n \sum x_{1t}x_{nt} &= \sum x_{1t}y_t \\ &\dots \dots \\ \beta_0 \sum x_{nt} + \beta_1 \sum x_{1t}x_{nt} + \dots + \beta_n \sum x_{nt}^2 &= \sum x_{nt}y_t \end{aligned}$$

求解上述方程组即可得到需求的回归系数 $\beta_0, \beta_1, \dots, \beta_n$ 。

2.3. 总方差估计及显著性检验

在多元线性回归模型中总方差的计算公式为：

$$S^2 = \frac{\sum e_i^2}{m-n}$$

其中, m 为样本总数, n 为回归系数, $\sum e_i^2$ 为残差平方和。且总方差 S^2 为随机误差项方差 σ^2 的无偏估计, 总方差越小表明误差越小, 回归方程越准确。

本文使用数学软件[3] IBM SPSS Statistics 对所求解的回归方程做显著性检验, 即分析方程的 R 平方和标准残差等数据。

3. 根据数据构建回归方程

将样本土地上的播种面积、有效灌溉面积、化肥施用量、机械总动力、成灾面积作为自变量, 该样本地区的粮食产量作为因变量, 从而构建相应的回归方程。

3.1. 构建回归方程

数据如表 1 所示:

Table 1. Food production data for 2008-2017

表 1. 2008 年~2017 年粮食产量数据

时间	粮食产量(万吨)	粮食作物播种面积(千公顷)	农业机械总动力(万千瓦)	有效灌溉面积(千公顷)	农用化肥施用折纯量(万吨)	受灾面积(千公顷)
2017 年	61,793.03	112,219.68	99,017.08	67,851	5859	18,478
2016 年	61,625.05	113,034.48	97,250.33	67,149	5984	26,221
2015 年	62,143.92	113,342.93	11,1728.1	65,872.6	6022.6	21,770
2014 年	60,702.61	112,722.58	108,056.6	64,540	5995.94	24,891
2013 年	60,193.84	111,955.56	103,906.75	63,473	5911.86	31,350
2012 年	58,957.97	111,204.59	102,558.96	62,490.5	5838.85	24,962
2011 年	57,120.85	110,573.02	97,734.66	61,682	5704.24	32,471
2010 年	54,647.71	109,876.09	92,780.48	60,348	5561.68	37,426
2009 年	53,082.08	108,985.75	87,496.1	59,261	5404.4	47,214
2008 年	52,870.92	106,792.65	82,190.41	58,472	5239	39,990

下面判断数据能否使用多元线性回归。根据多元线性回归方程的使用条件:

1) 如表 2 所示, 自变量对因变量有显著的影响, 并呈现出非确定关系(线性关系);

Table 2. Pearson relevance

表 2. Pearson 相关性

	粮食产量	粮食作物播种面积	农业机械总动力	有效灌溉面积	农用化肥施用折纯量	受灾面积
粮食产量	1	0.949	0.852	0.952	0.961	-0.925
粮食作物播种面积	0.949	1	0.892	0.899	0.985	-0.818
农业机械总动力	0.852	0.892	1	0.689	0.923	-0.779
有效灌溉面积	0.952	0.899	0.689	1	0.87	-0.889
农用化肥施用折纯量	0.961	0.985	0.923	0.87	1	-0.848
受灾面积	-0.925	-0.818	-0.779	-0.889	-0.848	1

- 2) 10 (样本容量) > 6 (系数个数);
- 3) 如图 1 所示, 残差服从正态分布。

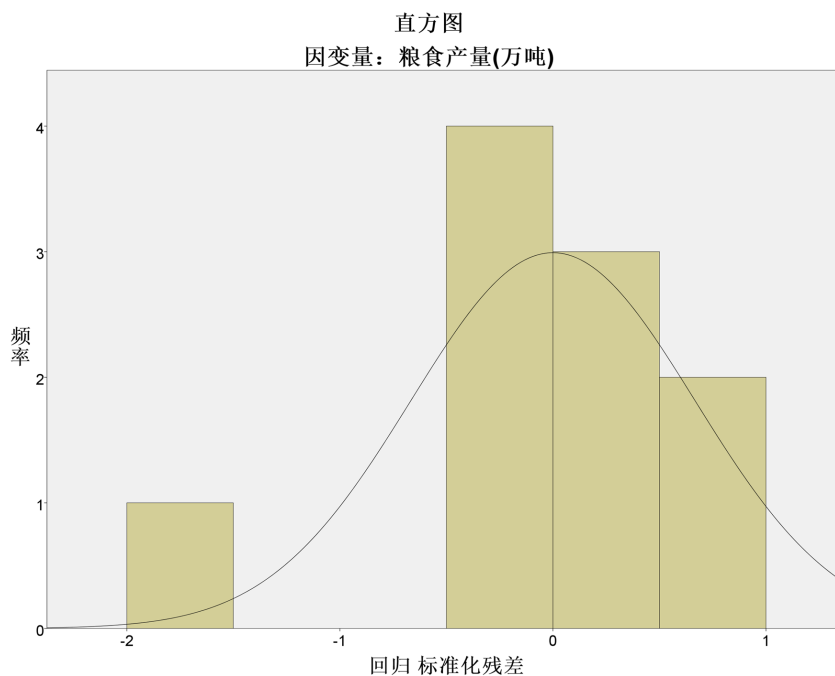


Figure 1. Standardized residual diagram
图 1. 标准化残差图

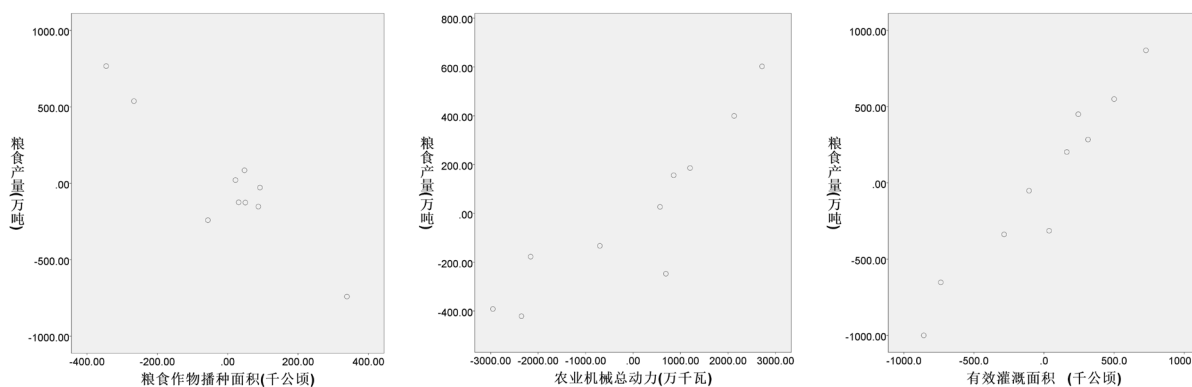
则可设五个自变量(粮食作物播种面积、农业机械总动力、有效灌溉面积、农用化肥施用折纯量和受灾面积)与因变量(粮食产量)的回归函数如下:

$$f(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \mu$$

其中, x_1 为粮食作物播种面积, x_2 为农业机械总动力, x_3 为有效灌溉面积, x_4 为农用化肥施用折纯量, x_5 为受灾面积, 且 $f(x)$ 为粮食产量, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ 为回归系数, μ 为误差, 数据量为十组(即 2008 年至 2017 年的六个指标, 来源于国家统计局)。

3.2. 求解回归方程

本文使用 IBM SPSS Statistics 对表 2 数据进行线性回归分析后, 绘制相关的偏回归图, 如图 2 所示:



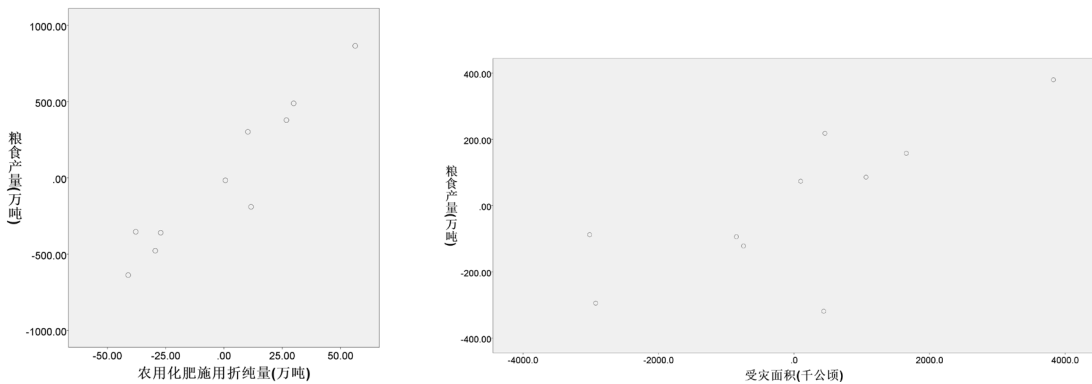


Figure 2. Partial regression diagrams of dependent and independent variables
图 2. 因变量与各自变量偏回归图

观察可得化肥施用量与粮食产量的相关性最高。

下用量化形式寻找上述五个变量与粮食产量之间的关系对表 1 中的数据进行线性回归后可解得 $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ ，即得到回归方程为：

$$f(x) = 114085.247 - 2.018x_1 + 0.155x_2 + 1.099x_3 + 14.134x_4 - 0.081x_5$$

可得到常数 β_0 的值为 114085.247，粮食作物播种面积系数 β_1 的值为 -2.018，农业机械总动力系数 β_2 的值为 0.155，有效灌溉面积系数 β_3 的值为 1.099，农用化肥施用折纯量 β_4 的值为 14.134，受灾面积系数 β_5 的值为 -0.081。即粮食产量与有效灌溉面积、农业机械总动力和农用化肥施用折纯量为正相关，与播种面积成负相关的主要原因为且与化肥施用量相关性最强，也符合科学的发展。

3.3. 显著性检验

根据 IBM SPSS Statistics 线性回归结果如表 3 所示：

Table 3. Test indicators and data values
表 3. 检验指标与数据值

检验指标	R 平方	残差平均值	标准残差
数据值	0.998319	-2.18E-11	0.666667

其标准残差为 0.667。其拟合程度 R^2 的值为 0.998，大于 0.95。即，此线性回归与模型的拟合程度良好。

4. 结合实际给出相关建议

总所周知，农业是一个国家至关重要的产业之一，而粮食生产量则是农业的重要指标之一，所以粮食产量的提高对国民经济的发展有重要作用。

可根据本文所得到的线性回归方程得到如下建议：

- 1) 通过线性回归方程可以看出，化肥施用量与粮食产量的正相关性最强，所以化肥可以有效促进粮食产量的提高。但在施用化肥的过程中要适量，否则会污染环境、降低土地肥力等，不利于长期发展。
- 2) 有效灌溉面积与粮食产量相关性较强，所以在粮食生产过程中，应有适量的水分，即在雨水不充足时采用人工降雨等措施。
- 3) 根据数据显示，播种面积与粮食产量成负相关，主要原因为生产技术的提高使得，即使是减小播

种面积也可以增加粮食的产量。即生产力的发展对粮食产量的提高有重大作用, 可节约农业用地, 使土地得到更合理的利用。所以应该合理的规划已有用地, 提高生产质量, 进而提高粮食产量。

4) 成灾面积与粮食产量成负相关, 即应尽量减少成灾面积, 减少粮食产量的损失。如及时给农作物驱赶害虫, 洪涝灾害时及时排水等。

参考文献

- [1] 李妍. 中国粮食生产因素及地区差异分析[J]. 经济研究导刊, 2009(13): 45-46.
- [2] 向东进, 李宏伟, 刘小雅. 实用多元统计分析[M]. 武汉: 中国地质大学(武汉)出版社, 2005.
- [3] 单良, 胡勇. 基于 Eviews, Excel, Spss 的回归分析比较[J]. 统计与决策, 2006(4): 150-152.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: aam@hanspub.org